

# Minimal entropy approximation for cellular automata

Henryk Fuks

*Department of Mathematics  
Brock University  
St. Catharines, Ontario L2S 3A1,  
Canada  
Email: hfuks@brocku.ca*

## Abstract

We present a method for construction of approximate orbits of measures under the action of cellular automata which is complementary to the local structure theory. The local structure theory is based on the idea of Bayesian extension, that is, construction of a probability measure consistent with given block probabilities and maximizing entropy. If instead of maximizing entropy one minimizes it, one can develop another method for construction of approximate orbits, at the heart of which is the iteration of finitely-dimensional maps, called minimal entropy maps. We present numerical evidence that minimal entropy approximation sometimes spectacularly outperforms the local structure theory in characterizing properties of cellular automata. Density response curve for elementary CA rule 26 is used to illustrate this claim.

## 1. Introduction

Let  $\mathcal{A} = \{0, 1, \dots, N-1\}$  be called an *alphabet*, or a *symbol set*, and let  $X = \mathcal{A}^{\mathbb{Z}}$ . Elements of  $X$  will be called configurations. A finite sequence of elements of  $\mathcal{A}$ ,  $\mathbf{b} = b_1 b_2 \dots b_n$  will be called a *block* (or *word*) of length  $n$ . Set of all blocks of elements of  $\mathcal{A}$  of all possible lengths will be denoted by  $\mathcal{A}^*$ . *Cylinder set* generated by the block  $\mathbf{b} = b_1 b_2 \dots b_n$  and anchored at  $i$  is defined as

$$[\mathbf{b}]_i = \{\mathbf{x} \in \mathcal{A}^{\mathbb{Z}} : \mathbf{x}_{[i, i+n)} = \mathbf{b}\}. \quad (1)$$

The appropriate mathematical description of a distribution of configurations is a probability measure on  $X$ . Cellular automata (CA) are often considered as maps in the space of such probability measures [1, 2, 3, 4].

In this paper, we will be interested in shift-invariant probability measures over  $X$ , or more precisely, in shift-invariant probability measures on the  $\sigma$ -algebra generated by elementary cylinder sets of  $X$ . Set of such measures will be denoted by  $\mathfrak{M}_{\sigma}(X)$ . Detailed construction

of measures in  $\mathfrak{M}_\sigma(X)$  is described in the review article [5], and interested reader is advised to consult this reference. Nevertheless, it is not necessary to be familiar with the details of the construction in order to follow the present paper.

The most important feature of any measure  $\mu \in \mathfrak{M}_\sigma(X)$  is that it is fully determined by measures of cylinder sets  $\mu([\mathbf{a}]_i)$  for all  $\mathbf{a} \in \mathcal{A}^*$ , which we will denote by

$$P(\mathbf{a}) = \mu([\mathbf{a}]_i). \quad (2)$$

Note that  $P(\mathbf{a})$ , which we will call *block probability*, is independent of  $i$  due to shift-invariance of the measure  $\mu$ . Block probability  $P(\mathbf{a})$  can be intuitively understood as the probability of occurrence of a given block  $\mathbf{a}$  in the distribution of configurations.

The following theorem formally states the connection between block probabilities and measures. It is a direct consequence of Hahn-Kolmogorov extension theorem. For proof the reader can consult [5] and references therein.

**Theorem 1.1** *Let  $P : \mathcal{A}^* \rightarrow [0, 1]$  satisfy the conditions*

$$P(\mathbf{b}) = \sum_{a \in \mathcal{A}} P(\mathbf{b}a) = \sum_{a \in \mathcal{G}} P(a\mathbf{b}) \quad \forall \mathbf{b} \in \mathcal{A}^*, \quad (3)$$

$$1 = \sum_{a \in \mathcal{G}} P(a). \quad (4)$$

*Then  $P$  uniquely determines shift-invariant probability measure on the  $\sigma$ -algebra generated by elementary cylinder sets of  $X$ .*

The conditions (3) and (4) are known in the literature as *consistency conditions*.

Although the set of all block probabilities  $\{P(\mathbf{a}) : \mathbf{a} \in \mathcal{A}^*\}$  is countable, it is still infinite, and in many practical problems, such as computer simulations, it is often possible to keep track of only a finite number of block probabilities. This brings an important question: if we know probabilities of all blocks of a given length, can we reconstruct the entire measure *approximately*? One answer to this question is well known and called “Bayesian extension”, originally introduced in the context of lattice gases [6, 7]. The approximate measure produced by the Bayesian extension is known as a “finite-block measure” or as “Markov process with memory”. The aforementioned review paper [5] discusses details of the Bayesian extension. The main feature of this extension is that, given a finite set of probabilities, it constructs all other block probabilities, satisfying consistency conditions, such that the resulting measure has the maximal entropy.

The Bayesian extension proved to be a very useful device in statistical physics as well as in the theory of cellular automata. In 1987, H. A. Gutowitz, J. D. Victor, and B. W. Knight [8] proposed a generalization of the mean-field theory for cellular automata based on the idea of Bayesian extension. They called it *local structure theory*. The local structure theory, recently formalized and extended [5], turned out to be a very powerful tool for characterization of cellular automata.

Given the success of the local structure theory, which is based on the maximal entropy approximation, it seems quite natural to ask how useful the complementary approximation would be, namely the one which *minimizes* the entropy instead of maximizing it? To the

knowledge of the author, no one has ever pursued this idea, and this article is intended to partially fill this gap.

In what follows, we will investigate the minimal entropy approximation in a configuration space over a binary alphabet, that is, assuming  $\mathcal{A} = \{0, 1\}$ . Although ideas presented below can be easily carried over to alphabets of higher cardinality, the binary case is the simplest and the most elegant one, and that is the only reason why we restrict our attention to  $\mathcal{A} = \{0, 1\}$ .

## 2. Minimal entropy extension

Before we proceed, let us define  $\mathbf{P}^{(k)}$  to be the column vector of all probabilities of blocks of length  $k$  arranged in lexical order. For example, for  $\mathcal{A} = \{0, 1\}$ , the first three vectors  $\mathbf{P}^{(k)}$  are

$$\begin{aligned}\mathbf{P}^{(1)} &= [P(0), P(1)]^T, \\ \mathbf{P}^{(2)} &= [P(00), P(01), P(10), P(11)]^T, \\ \mathbf{P}^{(3)} &= [P(000), P(001), P(010), P(011), P(100), P(101), P(110), P(111)]^T, \\ &\dots\end{aligned}$$

Entropy of  $\mathbf{P}^{(k)}$  will be defined as

$$h(\mathbf{P}^{(k)}) = - \sum_{\mathbf{b} \in \mathcal{A}^k} \hat{P}(\mathbf{b}) \log \hat{P}(\mathbf{b}). \quad (5)$$

Suppose that for a given probability measure we know all block probabilities  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(n)}$ . We want to construct block probabilities  $\mathbf{P}^{(n+1)}$  which minimize entropy  $h(\mathbf{P}^{(n+1)})$  and are consistent with block probabilities  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(n)}$ .

In order to do this, we first must remark that not all block probabilities which are elements of vectors  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(n)}$  are independent, due to consistency conditions. In [5], we demonstrated that for  $\mathcal{A} = \{0, 1\}$ , only  $2^{n-1}$  block probabilities are independent. Which ones are declared to be independent, and which ones are treated as dependent, is to some extent arbitrary. One choice of independent probabilities is called *short form representation* [5]. For the binary alphabet, in the short form representation block probabilities which have the form  $P(0\mathbf{a}0)$  are declared to be independent, and the remaining ones are treated as dependent. For example, among elements of  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \mathbf{P}^{(3)}$ , the independent probabilities

are  $P(0), P(00), P(000), P(010)$ . The remaining ones can be expressed as

$$\begin{aligned} \begin{bmatrix} P(001) \\ P(011) \\ P(100) \\ P(101) \\ P(110) \\ P(111) \end{bmatrix} &= \begin{bmatrix} P(00) - P(000) \\ P(0) - P(00) - P(010) \\ P(00) - P(000) \\ P(0) - 2P(00) + P(000) \\ P(0) - P(00) - P(010) \\ 1 - 3P(0) + 2P(00) + P(010) \end{bmatrix} \\ \begin{bmatrix} P(01) \\ P(10) \\ P(11) \end{bmatrix} &= \begin{bmatrix} P(0) - P(00) \\ P(0) - P(00) \\ 1 - 2P(0) + P(00) \end{bmatrix}, \\ P(1) &= 1 - P(0). \end{aligned} \tag{6}$$

Coming back to our problem, if we want to construct  $\mathbf{P}^{(n+1)}$  given  $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(n)}$ , we are free to choose only the values of elements of  $\mathbf{P}^{(n+1)}$  which are of the form  $P(0\mathbf{a}0)$ , where  $\mathbf{a} \in \mathcal{A}^{n-1}$ . These probabilities will be denoted by  $x_{\mathbf{a}}$ , and the remaining ones can be expressed in terms of  $x_{\mathbf{a}}$  and probabilities of shorter blocks,

$$\begin{aligned} P(0\mathbf{a}0) &= x_{\mathbf{a}}, \\ P(0\mathbf{a}1) &= P(0\mathbf{a}) - x_{\mathbf{a}}, \\ P(1\mathbf{a}0) &= P(\mathbf{a}0) - P(0\mathbf{a}0) = P(\mathbf{a}0) - x_{\mathbf{a}}, \\ P(1\mathbf{a}1) &= P(\mathbf{a}1) - P(0\mathbf{a}1) = P(\mathbf{a}1) - (P(0\mathbf{a}) - x_{\mathbf{a}}). \end{aligned} \tag{7}$$

The problem is now as follows: how to choose parameters  $x_{\mathbf{a}}$  in order to minimize entropy  $h(\mathbf{P}^{(n+1)})$ ? The following theorem provides the answer.

**Theorem 2.1** *Suppose that  $\mu$  is a shift-invariant probability measure, and  $P(\mathbf{b}) = \mu([\mathbf{b}]_i)$ . Let*

$$\begin{aligned} \hat{P}(0\mathbf{a}0) &= \hat{x}_{\mathbf{a}}, \\ \hat{P}(0\mathbf{a}1) &= P(0\mathbf{a}) - \hat{x}_{\mathbf{a}}, \\ \hat{P}(1\mathbf{a}0) &= P(\mathbf{a}0) - \hat{x}_{\mathbf{a}}, \\ \hat{P}(1\mathbf{a}1) &= P(\mathbf{a}1) - (P(0\mathbf{a}) - \hat{x}_{\mathbf{a}}), \end{aligned} \tag{8}$$

where

$$\hat{x}_{\mathbf{a}} = \begin{cases} \max \{0, P(0\mathbf{a}) - P(\mathbf{a}1)\} & \text{if } |P(\mathbf{a}1) - P(0\mathbf{a})| < |P(\mathbf{a}0) - P(0\mathbf{a})|, \\ \min \{P(0\mathbf{a}), P(\mathbf{a}0)\} & \text{otherwise.} \end{cases} \tag{9}$$

Then

$$- \sum_{\mathbf{b} \in \mathcal{A}^{n+1}} \hat{P}(\mathbf{b}) \log \hat{P}(\mathbf{b}) \leq - \sum_{\mathbf{b} \in \mathcal{A}^{n+1}} P(\mathbf{b}) \log P(\mathbf{b}). \tag{10}$$

*Proof.* Let us first notice that  $P(0\mathbf{a})$ ,  $P(\mathbf{a}0)$ , and  $P(\mathbf{a}1)$  are not independent. Consistency conditions imply that

$$P(0\mathbf{a}) \leq P(\mathbf{a}1) + P(\mathbf{a}0) \leq 1, \quad (11)$$

and from there we obtain

$$P(0\mathbf{a}) - P(\mathbf{a}0) \leq P(\mathbf{a}1) \leq 1 - P(\mathbf{a}0). \quad (12)$$

Denoting  $\alpha_{\mathbf{a}} = P(0\mathbf{a})$ ,  $\beta_{\mathbf{a}} = P(\mathbf{a}0)$ ,  $\delta_{\mathbf{a}} = P(\mathbf{a}1)$ , this can be written as

$$\alpha_{\mathbf{a}} - \beta_{\mathbf{a}} \leq \delta_{\mathbf{a}} \leq 1 - \beta_{\mathbf{a}}. \quad (13)$$

The right hand side of inequality (10) can be written as

$$\begin{aligned} & - \sum_{\mathbf{b} \in \mathcal{A}^{n+1}} P(\mathbf{b}) \log P(\mathbf{b}) = \\ & - \sum_{\mathbf{a} \in \mathcal{A}^{n-1}} P(0\mathbf{a}0) \log P(0\mathbf{a}0) + P(0\mathbf{a}1) \log P(0\mathbf{a}1) + P(1\mathbf{a}0) \log P(1\mathbf{a}0) + P(1\mathbf{a}1) \log P(1\mathbf{a}1). \end{aligned} \quad (14)$$

Using eqs. (7) this becomes

$$- \sum_{\mathbf{b} \in \mathcal{A}^{n+1}} P(\mathbf{b}) \log P(\mathbf{b}) = \sum_{\mathbf{a} \in \mathcal{A}^{n-1}} H_a(x_{\mathbf{a}}), \quad (15)$$

where we define

$$\begin{aligned} H_a(x_{\mathbf{a}}) = & -x_{\mathbf{a}} \log x_{\mathbf{a}} - (\alpha_{\mathbf{a}} - x_{\mathbf{a}}) \log (\alpha_{\mathbf{a}} - x_{\mathbf{a}}) \\ & -(\beta_{\mathbf{a}} - x_{\mathbf{a}}) \log (\beta_{\mathbf{a}} - x_{\mathbf{a}}) - (\delta_{\mathbf{a}} - (\alpha_{\mathbf{a}} - x_{\mathbf{a}})) \log (\delta_{\mathbf{a}} - (\alpha_{\mathbf{a}} - x_{\mathbf{a}})). \end{aligned} \quad (16)$$

Function  $H_a(x_{\mathbf{a}})$  is concave, and  $x_{\mathbf{a}}$  can only take values from some closed interval  $[x_{\mathbf{a},1}, x_{\mathbf{a},2}]$ . For this reason,  $H_a(x_{\mathbf{a}})$  reaches minimum at one of the endpoints of the interval. We will show that the minimum occurs precisely at  $x_{\mathbf{a}} = \hat{x}_{\mathbf{a}}$ , where  $\hat{x}_{\mathbf{a}}$  is defined in eq. (9). First, let us determine the values of the endpoints  $x_{\mathbf{a},1}, x_{\mathbf{a},2}$ . In order to do this, note that obviously  $x_{\mathbf{a}} \in [0, 1]$ . By consistency conditions,

$$P(\mathbf{a}1) = P(0\mathbf{a}1) + P(1\mathbf{a}1), \quad (17)$$

$$P(\mathbf{a}0) = P(0\mathbf{a}0) + P(1\mathbf{a}0), \quad (18)$$

$$P(0\mathbf{a}) = P(0\mathbf{a}0) + P(0\mathbf{a}1), \quad (19)$$

and therefore

$$P(\mathbf{a}1) \geq P(0\mathbf{a}1), \quad (20)$$

$$P(\mathbf{a}0) \geq P(0\mathbf{a}0), \quad (21)$$

$$P(0\mathbf{a}) \geq P(0\mathbf{a}0). \quad (22)$$

Using eqs. (7) and the notation  $\alpha_{\mathbf{a}} = P(0\mathbf{a})$ ,  $\beta_{\mathbf{a}} = P(\mathbf{a}0)$ ,  $\delta_{\mathbf{a}} = P(\mathbf{a}1)$ , this becomes

$$\delta_{\mathbf{a}} \geq \alpha_{\mathbf{a}} - x_{\mathbf{a}}, \quad (23)$$

$$\beta_{\mathbf{a}} \geq x_{\mathbf{a}}, \quad (24)$$

$$\alpha_{\mathbf{a}} \geq x_{\mathbf{a}}. \quad (25)$$

Solving the above system of inequalities for  $x_{\mathbf{a}}$  we obtain

$$\alpha_{\mathbf{a}} - \delta_{\mathbf{a}} \leq x_{\mathbf{a}} \leq \min \{\alpha_{\mathbf{a}}, \beta_{\mathbf{a}}\}. \quad (26)$$

Since  $x_{\mathbf{a}} \in [0, 1]$ , we obtain the following expression for the endpoints of the interval  $[x_{\mathbf{a},1}, x_{\mathbf{a},2}]$ ,

$$x_{\mathbf{a},1} = \max \{0, \alpha_{\mathbf{a}} - \delta_{\mathbf{a}}\}, \quad x_{\mathbf{a},2} = \min \{\alpha_{\mathbf{a}}, \beta_{\mathbf{a}}\}. \quad (27)$$

Suppose now that we fix both  $\mathbf{a}$  and  $\alpha_{\mathbf{a}}$ . Let us consider separately the four cases described in the table below.

	$\beta_{\mathbf{a}} < \alpha_{\mathbf{a}}$	$\beta_{\mathbf{a}} \geq \alpha_{\mathbf{a}}$
$\delta_{\mathbf{a}} < \alpha_{\mathbf{a}}$	$x_{\mathbf{a},1} = \alpha_{\mathbf{a}} - \delta_{\mathbf{a}}$ $x_{\mathbf{a},2} = \beta_{\mathbf{a}}$	$x_{\mathbf{a},1} = \alpha_{\mathbf{a}} - \delta_{\mathbf{a}}$ $x_{\mathbf{a},2} = \alpha_{\mathbf{a}}$
$\delta_{\mathbf{a}} \geq \alpha_{\mathbf{a}}$	$x_{\mathbf{a},1} = 0$ $x_{\mathbf{a},2} = \beta_{\mathbf{a}}$	$x_{\mathbf{a},1} = 0$ $x_{\mathbf{a},2} = \alpha_{\mathbf{a}}$

We will determine the sign of  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2})$ . If  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) < 0$ , then the minimum of  $H_{\mathbf{a}}$  occurs at  $x_{\mathbf{a},1}$ , otherwise at  $x_{\mathbf{a},2}$ . To avoid notational clutter, we will drop the index  $\mathbf{a}$  from  $\alpha_{\mathbf{a}}, \beta_{\mathbf{a}}, \delta_{\mathbf{a}}$ .

**Case 1:**  $\beta < \alpha, \delta < \alpha$ .

We have

$$H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) = \beta \log \beta + (\alpha - \beta) \log(\alpha - \beta) - (\alpha - \delta) \log(\alpha - \delta) - \delta \log \delta. \quad (28)$$

Defining  $f_p(x) = x \log x + (p - x) \log(p - x)$  we can write

$$H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) = f_{\alpha}(\beta) - f_{\alpha}(\delta). \quad (29)$$

The function  $f_{\alpha}(x)$ , defined on interval  $x \in (0, \alpha)$ , reaches minimum at  $x = \alpha/2$ , and has the property  $f_{\alpha}(x) = f_{\alpha}(\alpha - x)$ . This means that  $f_{\alpha}(\delta) > f_{\alpha}(\beta)$ , and thus  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) < 0$ , if and only if

$$\left| \delta - \frac{\alpha}{2} \right| > \left| \beta - \frac{\alpha}{2} \right|. \quad (30)$$

**Case 2:**  $\beta < \alpha, \delta \geq \alpha$ .

We have

$$\begin{aligned} H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) &= -\alpha \log \alpha - (\delta - \alpha) \log(\delta - \alpha) \\ &\quad + (\alpha - \beta) \log(\alpha - \beta) + (\beta - \alpha + \delta) \log(\beta - \alpha + \delta) = \\ &= f_{\delta}(\alpha - \beta) - f_{\delta}(\alpha). \end{aligned} \quad (31)$$

For the same reason as before,  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) < 0$  if and only if

$$\left| \alpha - \frac{\delta}{2} \right| > \left| \alpha - \beta - \frac{\delta}{2} \right|. \quad (32)$$

**Case 3:**  $\beta \geq \alpha, \delta < \alpha$ .

We have

$$\begin{aligned} H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) &= -(\alpha - \delta) \log(\alpha - \delta) \\ &\quad - (\beta - \alpha + \delta) \log(\beta - \alpha + \delta) + \alpha \log \alpha + (\beta - \alpha) \log(\beta - \alpha) \\ &= f_{\beta}(\alpha) - f_{\beta}(\alpha - \delta) \end{aligned} \quad (33)$$

Again, by the property of  $f_{\beta}$  discussed under Case 1,  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) < 0$  is equivalent to

$$\left| \alpha - \delta - \frac{\beta}{2} \right| > \left| \alpha - \frac{\beta}{2} \right|. \quad (34)$$

**Case 4:**  $\beta \geq \alpha, \delta \geq \alpha$ .

We have

$$\begin{aligned} H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) &= -\beta \log \beta - (\delta - \alpha) \log(\delta - \alpha) \\ &\quad + (\beta - \alpha) \log(\beta - \alpha) + \delta \log \delta \\ &= g(\delta) - g(\beta), \end{aligned} \quad (35)$$

where  $g(y) = y \log y - (y - \alpha) \log(y - \alpha)$ . Since for  $y \in (0, \alpha)$

$$g'(y) = \log y - \log(y - \alpha) > 0, \quad (36)$$

$h(y)$  is increasing in  $(0, \alpha)$ . This means that  $g(\delta) > g(\beta)$ , or equivalently  $H_{\mathbf{a}}(x_{\mathbf{a},1}) - H_{\mathbf{a}}(x_{\mathbf{a},2}) < 0$ , is satisfied if and only if

$$\delta < \beta. \quad (37)$$

We obtained four inequalities (30), (32), (34), and (37) for four cases. We plotted in Figure 1 solutions of these four inequalities in  $\beta$ - $\delta$  space using different colors for each case. One can, however, combine all four cases and describe them by one simple inequality taking into account the fact that only values of  $(\beta, \delta)$  marked by vertical hashing are possible, due to condition (13). This simple inequality combining all four cases (subject to condition (13)) is

$$|\delta - \alpha| < |\beta - \alpha|, \quad (38)$$

as one can easily verify graphically by comparing Figures 1 and 2.

To summarize our findings, we demonstrated that the minimum of  $H_{\mathbf{a}}$  occurs at  $\hat{x}_{\mathbf{a}}$ , where

$$\hat{x}_{\mathbf{a}} = \begin{cases} x_{\mathbf{a},1}, & \text{if } |\delta_{\mathbf{a}} - \alpha_{\mathbf{a}}| < |\beta_{\mathbf{a}} - \alpha_{\mathbf{a}}|, \\ x_{\mathbf{a},2} & \text{otherwise,} \end{cases} \quad (39)$$

and where  $x_{\mathbf{a},1}$  and  $x_{\mathbf{a},2}$  are defined in eq. (27). This is precisely eq. (9), and Theorem 2.1 then follows directly.  $\square$

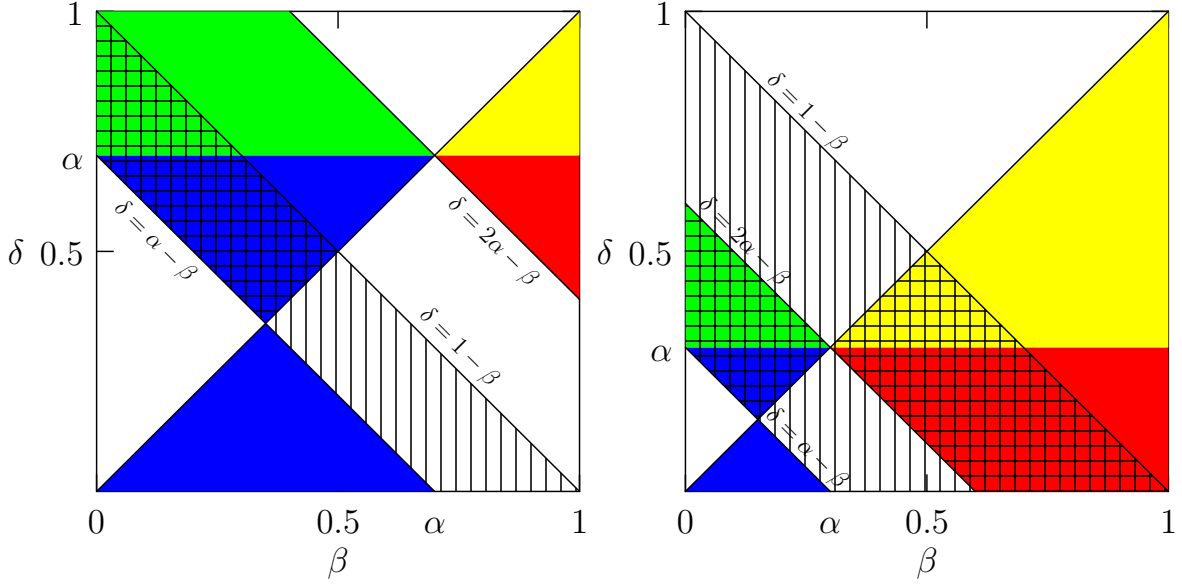


Figure 1: Solutions of inequalities (30), (32), (34) and (37) in  $\beta$ - $\delta$  space, shown, respectively, in blue, green, red, and yellow color. Region with vertical hatching represents solution of inequality (13), and the region with horizontal hatching represents parameters for which the minimum of  $H_a$  occurs at  $x_{a,1}$ . Two scenarios of are shown, corresponding to  $a > 0.5$  (left) and  $a < 0.5$  (right).

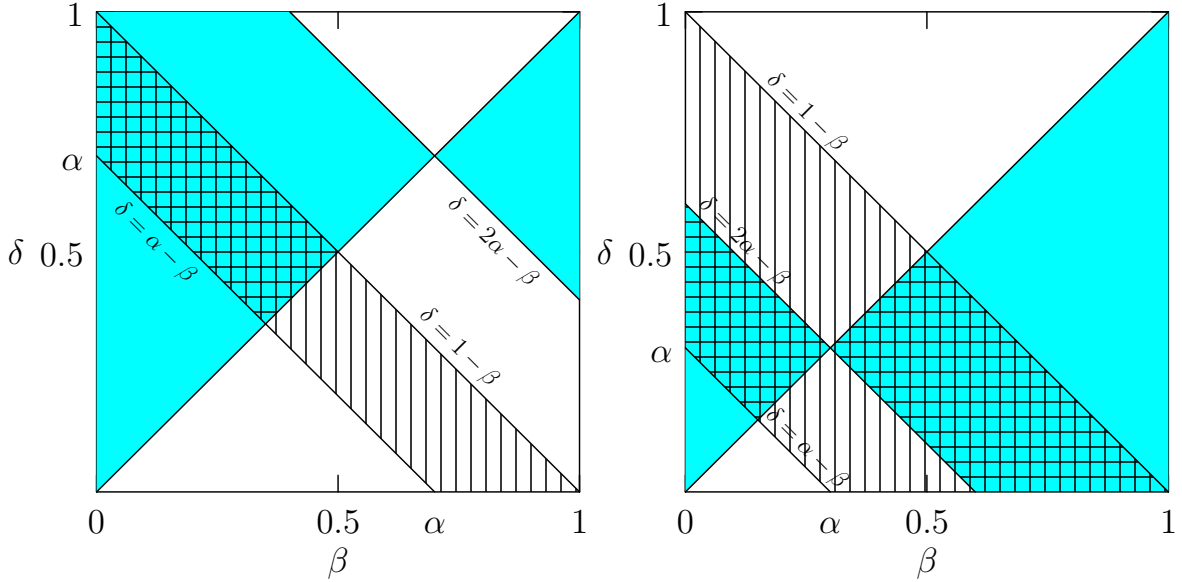


Figure 2: Simplification of Figure 1. Solutions of inequalities (30), (32), (34) and (37) have been replaced by solution of  $|\delta - \alpha| < |\beta - \alpha|$ , shown in cyan color.



### 3. Minimal entropy approximation for measures

Using Theorem 2.1, we can now construct approximation of a probability measure complementary to the Bayesian approximation. Let us define

$$\Upsilon(\alpha, \beta, \delta) = \begin{cases} \max\{0, \alpha - \delta\} & \text{if } |\delta - \alpha| < |\beta - \alpha|, \\ \min\{\alpha, \beta\} & \text{otherwise.} \end{cases} \quad (40)$$

and

$$\Upsilon_{0,0}(\alpha, \beta, \delta) = \Upsilon(\alpha, \beta, \delta), \quad (41)$$

$$\Upsilon_{0,1}(\alpha, \beta, \delta) = \alpha - \Upsilon(\alpha, \beta, \delta), \quad (42)$$

$$\Upsilon_{1,0}(\alpha, \beta, \delta) = \beta - \Upsilon(\alpha, \beta, \delta), \quad (43)$$

$$\Upsilon_{1,1}(\alpha, \beta, \delta) = \delta - \alpha + \Upsilon(\alpha, \beta, \delta). \quad (44)$$

Using this notation and eq. (8), one can now express probabilities of  $(k+1)$ -blocks by probabilities of  $k$ -blocks, by writing

$$P(a_1 a_2 \dots a_{k+1}) \approx \Upsilon_{a_1, a_{k+1}} \left( P(0 a_2 \dots a_k), P(a_2 \dots a_k 0), P(a_2 \dots a_k 1) \right). \quad (45)$$

The above approximation will be called *minimal entropy approximation*. We can, of course, repeat this process, and approximate  $(k+2)$ -block probabilities by  $(k+1)$ -block probabilities, and, by applying the approximation again,  $(k+1)$ -block probabilities by  $k$ -block probabilities. For a given integer  $k$ , by recursive application of the minimal entropy approximation, any block probability of length  $p > k$  can be expressed by probabilities of length  $k$ . The following proposition states this more formally.

**Proposition 3.1** *Let  $\mu \in \mathfrak{M}_\sigma(X)$  be a measure with associated block probabilities  $P : \mathcal{A}^* \rightarrow [0, 1]$ ,  $P(\mathbf{b}) = \mu([\mathbf{b}]_i)$  for all  $i \in \mathbb{Z}$  and  $\mathbf{b} \in \mathcal{A}^*$ . For  $k > 0$ , define  $\hat{P} : \mathcal{A}^* \rightarrow [0, 1]$  recursively so that*

$$\hat{P}(a_1 a_2 \dots a_p) = \begin{cases} P(a_1 a_2 \dots a_p) & \text{if } p \leq k, \\ \Upsilon_{a_1, a_p} \left( \hat{P}(0 a_2 \dots a_{p-1}), \hat{P}(a_2 \dots a_{p-1} 0), \hat{P}(a_2 \dots a_{p-1} 1) \right) & \text{if } p > k. \end{cases} \quad (46)$$

*Then  $\hat{P}$  determines a shift-invariant probability measure  $\hat{\mu}^{(k)} \in \mathfrak{M}_\sigma(X)$ , to be called minimal entropy approximation of  $\mu$  of order  $k$ .*

The proof that  $\hat{P}$  determines a measure is a direct consequence of the definition of  $\hat{P}$ .

It is intuitively clear that as the order of the minimal entropy approximations increases, the quality of the approximation should increase too. The following proposition formalizes this observation.

**Proposition 3.2** *The sequence of  $k$ -th order minimal entropy approximations of  $\mu \in \mathfrak{M}_\sigma(X)$  weakly converges to  $\mu$  as  $k \rightarrow \infty$ .*

*Proof:* Let  $n > 0$ ,  $\mathbf{b} \in \mathcal{A}^n$  and let  $\hat{P}_k(\mathbf{b}) = \hat{\mu}^{(k)}([\mathbf{b}]_0)$ ,  $P(\mathbf{b}) = \mu([\mathbf{b}]_0)$ . Since for  $k \geq n$   $\hat{P}_k(\mathbf{b}) = P(\mathbf{b})$ , we obviously have  $\lim_{k \rightarrow \infty} \hat{P}_k(\mathbf{b}) = P(\mathbf{b})$ . Since cylinder sets constitute convergence determining class for measures in  $\mathfrak{M}_\sigma(X)$ , convergence of block probabilities is equivalent to weak convergence. This leads to the conclusion that  $\hat{\mu}^{(k)} \Rightarrow \mu$ .  $\square$

Measures  $\mu$  for which  $\hat{\mu}^{(k)} = \mu$  will be called *k-th order measures of minimal entropy*. Set of such measures over  $X$  will be denoted by  $\mathfrak{M}_{\text{ME}}^{(k)}(X)$ . Obviously these measures are shift-invariant,  $\mathfrak{M}_{\text{ME}}^{(k)}(X) \subset \mathfrak{M}_\sigma^{(k)}(X)$ .

## 4. Orbits of measures under the action of cellular automata

Let  $w : \mathcal{A} \times \mathcal{A}^{2r+1} \rightarrow [0, 1]$ , whose values are denoted by  $w(a|\mathbf{b})$  for  $a \in \mathcal{A}$ ,  $\mathbf{b} \in \mathcal{A}^{2r+1}$ , satisfying  $\sum_{a \in \mathcal{A}} w(a|\mathbf{b}) = 1$ , be called *local transition function of radius r*, and its values will be called *local transition probabilities*. Probabilistic cellular automaton with local transition function  $w$  is a map  $F : \mathfrak{M}_\sigma(X) \rightarrow \mathfrak{M}_\sigma(X)$  defined as

$$(F\mu)([\mathbf{b}]_i) = \sum_{\mathbf{a} \in \mathcal{A}^{|\mathbf{b}|+2r}} w(\mathbf{a}|\mathbf{b}) \mu([\mathbf{a}]_{i-r}) \text{ for all } i \in \mathbb{Z}, \mathbf{b} \in \mathcal{A}^*, \quad (47)$$

where we define

$$w(\mathbf{a}|\mathbf{b}) = \prod_{j=1}^{|\mathbf{a}|} w(a_j | b_j b_{j+1} \dots b_{j+2r}). \quad (48)$$

When the function  $w$  takes values in the set  $\{0, 1\}$ , the corresponding cellular automaton is called *deterministic CA*.

For any probabilistic measure  $\mu \in \mathfrak{M}_\sigma(X)$ , we define the orbit of  $\mu$  under  $F$  as

$$\{F^n \mu\}_{n=0}^\infty. \quad (49)$$

Excluding trivial cases, computing the orbit of a measure under a given CA is very difficult, and no general method is known. We will, therefore, propose a method for approximating orbits based on the minimal entropy approximation.

Let us first define the *entropy minimizing operator* of order  $k$ , denoted by  $\Psi^{(k)}$ , to be a map from  $\mathfrak{M}_\sigma(X)$  to  $\mathfrak{M}_{\text{ME}}^{(k)}(X)$  such that

$$\Psi^{(k)} \mu = \hat{\mu}^{(k)}, \quad (50)$$

where  $\hat{\mu}^{(k)}$  is the measure defined in Proposition 3.1. Note that the operator  $\Psi^{(k)}$  is idempotent, that is,  $\Psi^{(k)} \Psi^{(k)} \mu = \Psi^{(k)} \mu$ . This allows us to construct approximate orbit of a measure  $\mu$  under the action of  $F$  by simply replacing  $F$  by  $\Psi^{(k)} F \Psi^{(k)}$ . The sequence

$$\left\{ \left( \Psi^{(k)} F \Psi^{(k)} \right)^n \mu \right\}_{n=0}^\infty \quad (51)$$

will be called the *minimal entropy approximation* of level  $k$  of the exact orbit  $\{F^n \mu\}_{n=0}^\infty$ . Note that all terms of this sequence are measures of minimal entropy, thus the entire approximate orbit lies in  $\mathfrak{M}_{\text{ME}}^{(k)}(X)$ .

Just like for the local structure approximation, the minimal entropy approximation approximates the actual orbit increasingly well as  $k$  increases. In fact, we will prove that every point of the approximate orbit weakly converges to the corresponding point of the exact orbit.

**Proposition 4.1** *Let  $k$  be a positive integer and  $\mathbf{b} \in \mathcal{A}^*$ . If  $k \geq |\mathbf{b}| + 2r$ , then*

$$F\mu([\mathbf{b}]) = F\Psi^{(k)}\mu([\mathbf{b}]) = \Psi^{(k)}F\mu([\mathbf{b}]). \quad (52)$$

*Proof.* To prove it, note that  $\mu([\mathbf{a}]) = \hat{\mu}^{(k)}([\mathbf{a}])$  for all blocks  $\mathbf{a}$  of length up to  $k$ . The first equality of (52) can be written as

$$\sum_{\mathbf{a} \in \mathcal{A}^{|\mathbf{b}|+2r}} w(\mathbf{a}|\mathbf{b})\mu([\mathbf{a}]) = \sum_{\mathbf{a} \in \mathcal{A}^{|\mathbf{b}|+2r}} w(\mathbf{a}|\mathbf{b})\hat{\mu}^{(k)}([\mathbf{a}]). \quad (53)$$

The equality holds when  $|\mathbf{a}| \leq k$ , that is,  $|\mathbf{b}| + 2r \leq k$ .

The second equality of (52) is a result of the fact that the  $\Psi^{(k)}$  operator only modifies probabilities of blocks of length greater than  $k$ . Since  $k \geq |\mathbf{b}| + 2r$ , we have  $|\mathbf{b}| < k$  and therefore  $F\mu([\mathbf{b}]) = \Psi^{(k)}F\mu([\mathbf{b}])$ .  $\square$

Now let us note that  $F^n$  can be viewed as a cellular automaton rule of radius  $nr$ , thus when  $k \geq |\mathbf{b}| + 2nr$ , we have  $F^n\mu([\mathbf{b}]) = F^n\Psi^{(k)}\mu([\mathbf{b}])$ . We can insert arbitrary number of  $\Psi^{(k)}$  operators on the right hand side anywhere we want, and nothing will change, because  $\Psi^{(k)}$  does not modify relevant block probabilities. This yields an immediate corollary.

**Corollary 4.1** *Let  $k$  and  $n$  be positive integers and  $\mathbf{b} \in \mathcal{A}^*$ . If  $k \geq |\mathbf{b}| + 2nr$ , then*

$$F^n\mu([\mathbf{b}]) = (\Psi^{(k)}F\Psi^{(k)})^n\mu([\mathbf{b}]).$$

This means that for a given  $n$ , measures of cylinder sets in the approximate measure  $(\Psi^{(k)}F\Psi^{(k)})^n\mu$  coincide with measures of cylinder sets in  $F^n\mu$  for sufficiently large  $k$ . Because cylinder sets constitute convergence determining class for measures, we obtain the following result.

**Theorem 4.1** *Let  $F$  be a cellular automaton,  $\mu \in \mathfrak{M}_\sigma(X)$  be a shift-invariant measure, and  $\nu_n^{(k)}$  be a minimal entropy approximation of order  $k$  of the measure  $F^n\mu$ , i.e.,  $\nu_n^{(k)} = (\Psi^{(k)}F\Psi^{(k)})^n\mu$ . Then for any positive integer  $n$ ,  $\nu_n^{(k)} \Rightarrow F^n\mu$  as  $k \rightarrow \infty$ .*

## 5. Minimal entropy maps

Minimal entropy measures can be entirely described by specifying a finite number of block probabilities. We will use this feature to construct a finite-dimensional map which approximates the action of a CA rule on shift-invariant measures. If  $\nu_n^{(k)} = (\Psi^{(k)}F\Psi^{(k)})^n\mu$ , then  $\nu_n^{(k)}$  satisfies recurrence equation

$$\nu_{n+1}^{(k)} = \Psi^{(k)}F\Psi^{(k)}\nu_n^{(k)}. \quad (54)$$

On both sides of this equation we have measures in  $\mathfrak{M}_{\text{ME}}^{(k)}(X)$ , and these are completely determined by probabilities of blocks of length  $k$ . If  $|\mathbf{b}| = k$ , we obtain

$$\nu_{n+1}^{(k)}([\mathbf{b}]) = \Psi^{(k)} F \Psi^{(k)} \nu_n^{(k)}([\mathbf{b}]), \quad (55)$$

and, since  $\Psi^{(k)}$  does not modify probabilities of blocks of length  $k$ , this simplifies to

$$\nu_{n+1}^{(k)}([\mathbf{b}]) = F \Psi^{(k)} \nu_n^{(k)}([\mathbf{b}]). \quad (56)$$

By the definition of  $F$ ,

$$\nu_{n+1}^{(k)}([\mathbf{b}]) = \sum_{\mathbf{a} \in \mathcal{A}^{|\mathbf{b}|+2r}} w(\mathbf{a}|\mathbf{b}) (\Psi^{(k)} \nu_n^{(k)})([\mathbf{a}]). \quad (57)$$

To simplify the notation, let us define  $Q_n(\mathbf{b}) = \nu_n^{(k)}([\mathbf{b}])$ , and, consistent with definition in eq. (46),  $\widehat{Q}_n(\mathbf{a}) = (\Psi^{(k)} \nu_n^{(k)})([\mathbf{a}])$ . Then we can rewrite the previous equation to take the form

$$Q_{n+1}(\mathbf{b}) = \sum_{\mathbf{a} \in \mathcal{A}^{|\mathbf{b}|+2r}} w(\mathbf{a}|\mathbf{b}) \widehat{Q}_n(\mathbf{a}). \quad (58)$$

Note that by eq. (46),  $\widehat{Q}_n(\mathbf{a})$  depends only on probabilities of blocks of length  $k$ . If we thus arrange  $Q_n(\mathbf{b})$  for all  $\mathbf{b} \in \mathcal{A}^k$  in lexicographical order to form a vector  $\mathbf{Q}_n$ , we will obtain

$$\mathbf{Q}_{n+1} = U^{(k)}(\mathbf{Q}_n), \quad (59)$$

where  $U^{(k)} : [0, 1]^{|\mathcal{A}|^k} \rightarrow [0, 1]^{|\mathcal{A}|^k}$  has components defined by eq. (58). We will call this map *an entropy minimizing map of order  $k$* .

## 6. Example: elementary CA rule 26

As an example, consider rule 26 given by

$$\begin{aligned} w(1|000) &= 0, w(1|001) = 1, w(1|010) = 0, w(1|011) = 1, \\ w(1|100) &= 1, w(1|101) = 0, w(1|110) = 0, w(1|111) = 0, \end{aligned} \quad (60)$$

and suppose we wish to construct minimal entropy map of order 2 for this rule. Let  $P_n(\mathbf{b}) = F^n \mu([\mathbf{b}])$ . Using eq. (47) we obtain for  $r = 1$ ,  $|\mathbf{b}| = 3$

$$P_{n+1}(\mathbf{b}) = \sum_{\mathbf{a} \in \mathcal{A}^5} w(\mathbf{a}|\mathbf{b}) P_n(\mathbf{a}). \quad (61)$$

Using definition of  $w(\mathbf{a}|\mathbf{b})$  given in eq. (48) and transition probabilities given in eq. (60) we obtain

$$\begin{aligned} P_{n+1}(00) &= P_n(0000) + P_n(0101) + P_n(1010) + P_n(1101) + P_n(1110) + P_n(1111), \\ P_{n+1}(01) &= P_n(0001) + P_n(0100) + P_n(1011) + P_n(1100), \\ P_{n+1}(10) &= P_n(0010) + P_n(0110) + P_n(0111) + P_n(1000), \\ P_{n+1}(11) &= P_n(0011) + P_n(1001). \end{aligned} \quad (62)$$

This set of equations describes exact relationship between block probabilities at step  $n + 1$  and block probabilities at step  $n$ . Note that 3-block probabilities at step  $n + 1$  are given in terms of 5-blocks probabilities at step  $n$ , thus it is not possible to iterate these equations.

Minimal entropy map of order 2 (eq. 59) can be obtained by simply replacing  $P$  by  $Q$  and placing the operator  $\hat{\cdot}$  over probabilities on the right hand side of eq. (62). This yields

$$\begin{aligned} Q_{n+1}(00) &= \hat{Q}_n(0000) + \hat{Q}_n(0101) + \hat{Q}_n(1010) + \hat{Q}_n(1101) + \hat{Q}_n(1110) + \hat{Q}_n(1111), \\ Q_{n+1}(01) &= \hat{Q}_n(0001) + \hat{Q}_n(0100) + \hat{Q}_n(1011) + \hat{Q}_n(1100), \\ Q_{n+1}(10) &= \hat{Q}_n(0010) + \hat{Q}_n(0110) + \hat{Q}_n(0111) + \hat{Q}_n(1000), \\ Q_{n+1}(11) &= \hat{Q}_n(0011) + \hat{Q}_n(1001). \end{aligned} \quad (63)$$

Using eq. (46) with  $k = 2$ , one can express  $\hat{Q}_n(a_1 a_2 a_3 a_4)$  in terms of 2-block probabilities. For example,

$$\begin{aligned} \hat{Q}_n(0000) &= \Upsilon(\hat{Q}_n(000), \hat{Q}_n(000), \hat{Q}_n(001)) = \hat{Q}_n(000) \\ &= \Upsilon(\hat{Q}_n(00), \hat{Q}_n(00), \hat{Q}_n(01)) = \hat{Q}_n(00) = Q_n(00). \end{aligned} \quad (64)$$

Similarly one obtains

$$\hat{Q}_n(0101) = Q_n(01), \quad \hat{Q}_n(1010) = Q_n(10), \quad \hat{Q}_n(1111) = Q_n(11). \quad (65)$$

All other  $\hat{Q}_n(a_1 a_2 a_3 a_4)$  are equal to 0. This simplifies eq. (63) to

$$\begin{aligned} Q_{n+1}(00) &= Q_n(00) + Q_n(01) + Q_n(10) + Q_n(11), \\ Q_{n+1}(01) &= 0, \\ Q_{n+1}(10) &= 0, \\ Q_{n+1}(11) &= 0. \end{aligned} \quad (66)$$

This defines a minimal entropy map  $U^{(k)} : [0, 1]^4 \rightarrow [0, 1]^4$  (cf. eq. 59) which can be iterated, albeit in this case, it is a trivial map, which after one iteration reaches the fixed point  $(1, 0, 0, 0)$ , because  $Q_n(00) + Q_n(01) + Q_n(10) + Q_n(11) = 1$ . We need higher order approximation in order to obtain a more “interesting” map.

When  $k = 3$ , we follow the same procedure as for the  $k = 2$  case discussed above. If we write eq. (58) for all possible  $\mathbf{b} \in \mathcal{A}^3$ , we will have on the left hand sides eight block

probabilities  $Q(b_1b_2b_3)$ , thus the resulting minimal entropy map will be 8-dimensional,

$$\begin{aligned}
Q_{n+1}(000) &= \widehat{Q}_n(00000) + \widehat{Q}_n(01010) + \widehat{Q}_n(10101) + \widehat{Q}_n(11010) + \widehat{Q}_n(11101) \\
&\quad + \widehat{Q}_n(11110) + \widehat{Q}_n(11111), \\
Q_{n+1}(001) &= \widehat{Q}_n(00001) + \widehat{Q}_n(01011) + \widehat{Q}_n(10100) + \widehat{Q}_n(11011) + \widehat{Q}_n(11100), \\
Q_{n+1}(010) &= \widehat{Q}_n(00010) + \widehat{Q}_n(01000) + \widehat{Q}_n(10110) + \widehat{Q}_n(10111) + \widehat{Q}_n(11000), \\
Q_{n+1}(011) &= \widehat{Q}_n(00011) + \widehat{Q}_n(01001) + \widehat{Q}_n(11001), \\
Q_{n+1}(100) &= \widehat{Q}_n(00101) + \widehat{Q}_n(01101) + \widehat{Q}_n(01110) + \widehat{Q}_n(01111) + \widehat{Q}_n(10000), \\
Q_{n+1}(101) &= \widehat{Q}_n(00100) + \widehat{Q}_n(01100) + \widehat{Q}_n(10001), \\
Q_{n+1}(110) &= \widehat{Q}_n(00110) + \widehat{Q}_n(00111) + \widehat{Q}_n(10010), \\
Q_{n+1}(111) &= \widehat{Q}_n(10011).
\end{aligned} \tag{67}$$

On the right hand side, we have 32 block probabilities which have to be expressed in terms of 3-block probabilities by using eq. (46) with  $k = 3$ . Some of these will simplify to a single 3-block probability, e.g.,

$$\widehat{Q}_n(00000) = \Upsilon(\widehat{Q}_n(0000), \widehat{Q}_n(0000), \widehat{Q}_n(0001)) = \widehat{Q}_n(0000) = Q_n(000). \tag{68}$$

Others, in general, will not simplify, and will have to be expressed by nested  $\Upsilon$  functions, for example

$$\begin{aligned}
\widehat{Q}_n(00100) &= \Upsilon\left(\Upsilon(Q_n(001), Q_n(010), Q_n(011)), \Upsilon(Q_n(010), Q_n(100), Q_n(101)), \right. \\
&\quad \left. \Upsilon_{0,1}(Q_n(010), Q_n(100), Q_n(101))\right). \tag{69}
\end{aligned}$$

Once we express all  $\widehat{Q}_n(a_1a_2a_3a_4a_5)$  in eq. (67) by 3-block probabilities  $Q_n(a_1a_2a_3)$ , we obtain a map  $[0, 1]^8 \rightarrow [0, 1]^8$ . We omit explicit formulae for this map due to its complexity. One should stress, however, that only four components of this map are independent, and that by exploiting consistency conditions for block probabilities it is possible to reduce this map to  $[0, 1]^4 \rightarrow [0, 1]^4$ . We refer interested reader to [5], where we explained how to perform such reduction for local structure maps (the same method can be used for minimal entropy maps).

Just for the sake of comparison, let us also write local structure map of order 3 for rule 26.

It can be obtained from eq. (67) by replacing  $\widehat{Q}$  with  $\widetilde{Q}$ ,

$$\begin{aligned}
Q_{n+1}(000) &= \widetilde{Q}_n(00000) + \widetilde{Q}_n(01010) + \widetilde{Q}_n(10101) + \widetilde{Q}_n(11010) + \widetilde{Q}_n(11101) \\
&\quad + \widetilde{Q}_n(11110) + \widetilde{Q}_n(11111), \\
Q_{n+1}(001) &= \widetilde{Q}_n(00001) + \widetilde{Q}_n(01011) + \widetilde{Q}_n(10100) + \widetilde{Q}_n(11011) + \widetilde{Q}_n(11100), \\
Q_{n+1}(010) &= \widetilde{Q}_n(00010) + \widetilde{Q}_n(01000) + \widetilde{Q}_n(10110) + \widetilde{Q}_n(10111) + \widetilde{Q}_n(11000), \\
Q_{n+1}(011) &= \widetilde{Q}_n(00011) + \widetilde{Q}_n(01001) + \widetilde{Q}_n(11001), \\
Q_{n+1}(100) &= \widetilde{Q}_n(00101) + \widetilde{Q}_n(01101) + \widetilde{Q}_n(01110) + \widetilde{Q}_n(01111) + \widetilde{Q}_n(10000), \\
Q_{n+1}(101) &= \widetilde{Q}_n(00100) + \widetilde{Q}_n(01100) + \widetilde{Q}_n(10001), \\
Q_{n+1}(110) &= \widetilde{Q}_n(00110) + \widetilde{Q}_n(00111) + \widetilde{Q}_n(10010), \\
Q_{n+1}(111) &= \widetilde{Q}_n(10011),
\end{aligned} \tag{70}$$

where

$$\widetilde{Q}_n(a_1 a_2 a_3 a_4 a_5) = \frac{Q_n(a_1 a_2 a_3) Q_n(a_2 a_3 a_4) Q_n(a_3 a_4 a_5)}{(Q_n(a_2 a_3 0) + Q_n(a_2 a_3 1))(Q_n(a_3 a_4 0) + Q_n(a_3 a_4 1))}. \tag{71}$$

Both minimal entropy maps and local structure maps become rather complicated when  $k$  increases. Because of high dimensionality and strong nonlinearity, it is difficult to perform standard stability analysis for these maps. It is, however, rather straightforward to write a computer program which constructs and iterates them.

## 7. Experimental results

As we already mentioned, orbits of minimal entropy maps approximate orbits of measures under cellular automata rules. By iterating the minimal entropy map, we can obtain approximate  $P_n(\mathbf{a})$ , that is, probability of occurrence of block  $\mathbf{a}$  after  $n$  iterations of a given cellular automata rule. How good is this approximation, and it is any better than the local structure approximation?

In order to shed some light on this question, we considered the following problem. Suppose that the initial measure is a Bernoulli measure  $\mu_p$ , so that

$$\mu_p([\mathbf{a}]) = P_0(\mathbf{a}) = p^j (1-p)^{|\mathbf{a}|-j}, \tag{72}$$

where  $j$  is the number of ones in  $\mathbf{a}$ ,  $|\mathbf{a}| - j$  is the number of zeros in  $\mathbf{a}$ , and  $p \in [0, 1]$ . Probability of occurrence of  $\mathbf{a}$  after  $n$  iterations is then given by

$$P_n(\mathbf{a}) = (F^n \mu_p)([\mathbf{a}]). \tag{73}$$

The expected value of a given cell after  $n$ -th iteration of the rule, to be denoted  $\rho_n$ , is given by

$$\rho_n = 1 \cdot P_n(1) + 0 \cdot P_n(0) = P_n(1). \tag{74}$$

We will call  $\rho_n$  a *density of ones* at time  $n$ . Density can be estimated numerically by starting with an array of  $N$  sites and setting each one of them independently to 1 or 0 with probability

$p$  or  $1-p$ , respectively. We then iterate rule  $F$   $n$  times (using periodic boundary conditions) and count how many cells are in state 1. The count divided by  $N$  serves as a numerical estimate of  $\rho_n$ .

One can also estimate  $\rho_n$  by iterating  $k$ -th order minimal entropy map  $n$  times starting from initial conditions given by eq. (72), that is,  $Q_0(\mathbf{a}) = P_0(\mathbf{a})$ . Then we compute  $Q_n(1)$  by using consistency conditions,

$$Q_n(1) = \sum_{\mathbf{a} \in \{0,1\}^{k-1}} Q_n(\mathbf{a}1), \quad (75)$$

and  $Q_n(1)$  is used as an approximation of  $\rho_n$ , to be called  $k$ -th order minimal entropy approximation of  $\rho_n$ . Analogous approximation using local structure map will be called  $k$ -th order local structure approximation of  $\rho_n$ .

An interesting question is now how  $\rho_n$  depends on  $\rho_0$ . Plot of  $\rho_n$  vs.  $\rho_0$  is called *density response curve*. We plotted density response curves using “experimental”  $\rho_n$  as well as using minimal entropy approximation and local structure approximation, both for orders  $k = 1, 2, \dots, 7$ . We found that, generally, as the order of the approximation increases, density response curves obtained by iterating minimal entropy maps become closer and closer to “experimental curves”. The same phenomenon is observed for density response curves obtained by iterating local structure maps.

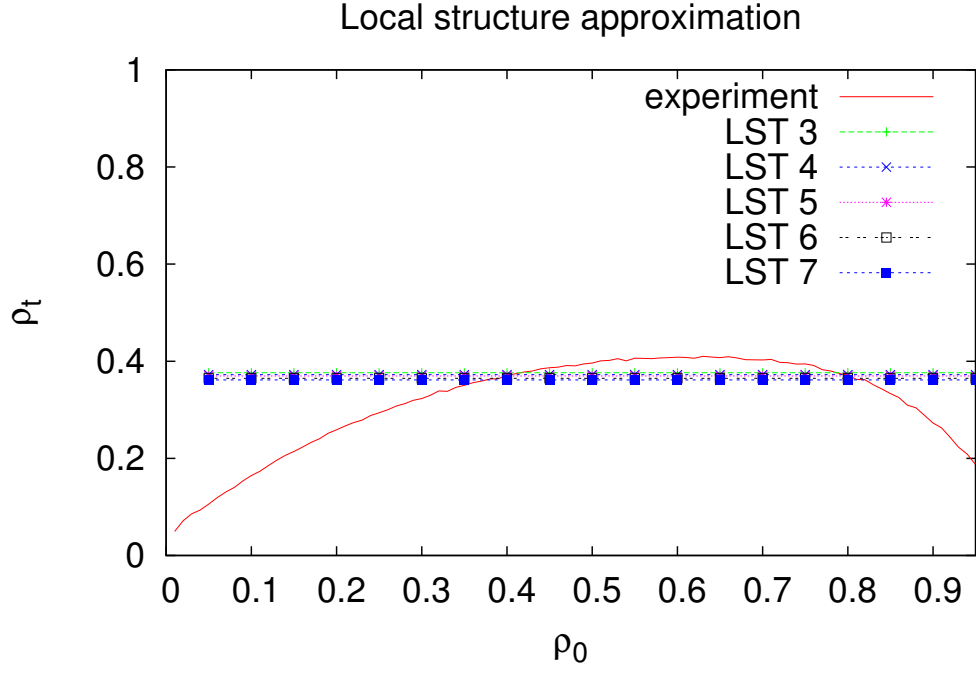
For most elementary rules, both local structure maps and minimal entropy maps produce good approximations of density curves. There are two exceptions, however, elementary CA rules 26 and 41. Here we will discuss rule 26 as an example. The experimental density response curve is shown as the continuous curve in Figure 3. Remarkably, density curves obtained by iterations of local structure maps up to order 7 are horizontal straight lines, as shown in Figure 3(a). One can say, therefore, that the local structure fails to predict the correct shape of the density curve, at least for  $k \leq 7$ .

In contrast to this, density curves obtained by iterations of minimal entropy maps, shown in Figure 3(a), approximate the shape of the “experimental” density curve much better, even at order 3. The minimal entropy approximation, therefore, clearly outperforms the local structure approximation in this case.

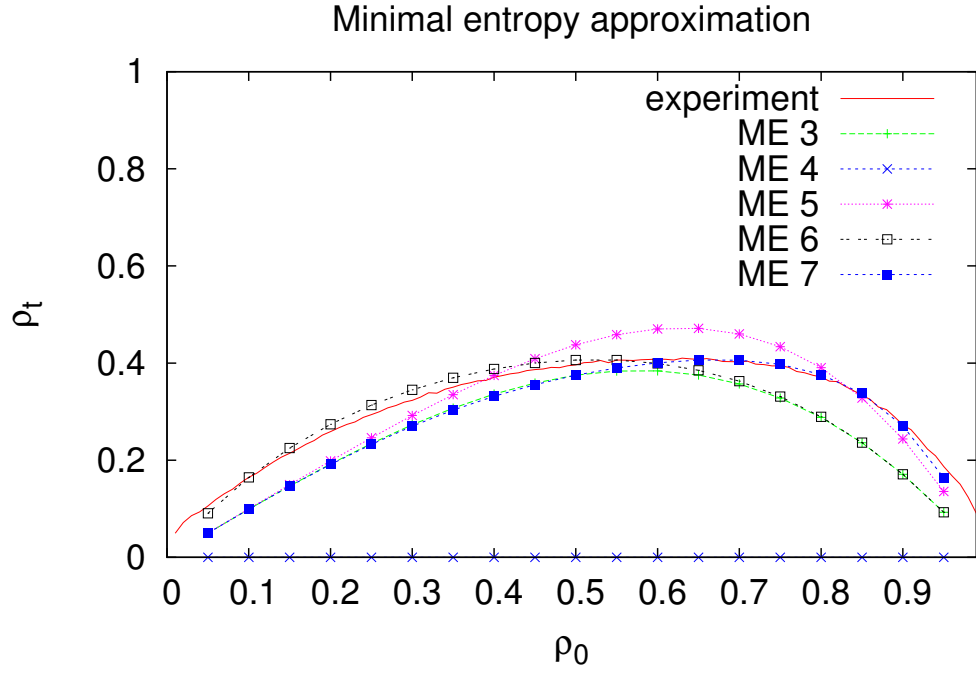
## 8. Conclusions

We introduced the notion of the minimal entropy approximation of probability measures over binary bisequences. Minimal entropy approximation can be viewed as an opposite of Bayesian approximation, which maximizes entropy. We then demonstrated how the minimal entropy approximation can be used to construct approximations of orbits of measures under the action of deterministic or probabilistic cellular automata. Such approximate orbits can be fully characterized by orbits of finite-dimensional maps, which we call minimal entropy maps. While points of approximate orbits of measures obtained by iterating minimal entropy maps weakly converge to corresponding points of the exact orbits, just as in the case of approximate orbits of local structure theory, there are cases when the minimal entropy approximation works better than the local structure approximation. This is the case for





(a)



(b)

Figure 3: Density response curves for rule 26 for  $t = 10^3$  obtained by iteration of local structure maps (a) and minimal entropy maps (b).

elementary CA rule 26, for which the local structure theory fails in predicting the correct shape of the density response curve for  $k \leq 7$ . The minimal entropy approximation yields fairly accurate prediction for the density response curve of rule 26, starting with  $k = 3$ .

An interesting question is why is the minimal entropy approximation better than the maximal entropy approximation in the case of rule 26? One could naively think that this is because the time evolution of rule 26 is somewhat more “ordered” than for other rules. It is, however, not true: there are other rules for which the spatiotemporal patterns are even more “ordered” than for rule 26, yet both maximal and minimal entropy approximations seem to work for them equally well. In order to probe this issue further, one will need to find more examples of rules for which the minimal entropy approximation outperforms the local structure theory. A natural way to go beyond elementary CA rules considered here is to search for such examples among either probabilistic CA rules of radius 1, or deterministic CA rules of radius greater than 1. Both possibilities are currently investigated by the author.

## 9. Acknowledgements

The author acknowledges partial financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of Discovery Grant. Some calculations on which this work is based were made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.

## References

- [1] P. Kůrka and A. Maass, “Limit sets of cellular automata associated to probability measures,” *Journal of Statistical Physics* **100** (2000) 1031–1047.
- [2] P. Kůrka, “On the measure attractor of a cellular automaton,” *Discrete and Continuous Dynamical Systems* (2005) 524 – 535.
- [3] M. Pivato, “Ergodic theory of cellular automata,” in *Encyclopedia of Complexity and System Science*, R. A. Meyers, ed. Springer, 2009.
- [4] E. Formenti and P. Kůrka, “Dynamics of cellular automata in non-compact spaces,” in *Encyclopedia of Complexity and System Science*, R. A. Meyers, ed. Springer, 2009.
- [5] H. Fukś, “Construction of local structure maps for cellular automata,” *J. of Cellular Automata* **7** (2013) 455–488, [arXiv:1304.8035](#).
- [6] H. J. Brascamp, “Equilibrium states for a one dimensional lattice gas,” *Communications In Mathematical Physics* **21** (1971), no. 1, 56.
- [7] M. Fannes and A. Verbeure, “On solvable models in classical lattice systems,” *Commun. Math. Phys.* **96** (1984) 115–124.

- [8] H. A. Gutowitz, J. D. Victor, and B. W. Knight, “Local structure theory for cellular automata,” *Physica D* **28** (1987) 18–48.